Volume 13 Number 22 1985 Nucleic Acids Research

DNA sequence of the region in the genome of herpes simplex virus type 1 containing the genes for DNA polymerase and the major DNA binding protein

John P.Quinn+ and Duncan J.McGeoch*

Institute of Virology, Church Street, Glasgow G11 5JR, UK

Received 4 October 1985; Accepted 24 October 1985

ABSTRACT

In the long unique region of the genome of herpes simplex virus type 1 (HSV-1), the genes for DNA polymerase and the major DNA binding protein are arranged in a head to head manner, with an origin of DNA replication (termed Ori_L) located between them. This paper reports an 8400 base pair DNA sequence containing both genes and the origin, obtained mostly by M13/dideoxy analysis of plasmid cloned fragments. Amino acid sequences of the two proteins were deduced. Homologues of both genes were detected in the genome sequence of the distantly related Epstein-Barr virus (EBV). Arrangement of these HSV-1 and EBV genes differs in genome location and in relative orientation. A part of HSV-1 DNA polymerase was found to be similar to a sequence in adenovirus 2 DNA polymerase, but the significance of this is unclear. Since a DNA sequence in the locality of Ori_L deletes on plasmid cloning, this region was analysed using virus DNA. A palindrome with 72-residue arms was found, which shows great similarity to the better characterized origin, Ori_S.

INTRODUCTION

In 1963, Keir and Gold (1) described the induction of DNA dependent DNA polymerase in culture cells infected by herpes simplex virus (HSV). This work marked the start of enzymatic study of herpesvirus DNA replication. Since then the DNA polymerase has been extensively studied, and it is now clear that it is encoded in the HSV genome, and that it participates in replication of the genome (2). HSV DNA polymerase has been purified and shown to be a single polypeptide chain of molecular weight about 140,000 (3,4). In addition to DNA polymerising activity, the protein also exhibits a 3' to 5' exonuclease action, thought to represent a proofreading function (3). The DNA polymerase has also emerged as a key component in the mechanism of action of several antiviral compounds (5-9). The gene for HSV DNA polymerase (here called pol) has been located in

the long unique region of the genome (U_L: see Figure 1) near 0.4 map units (6). *Ts*, drug resistance and drug hypersensitivity mutants in the polymerase gene have been described (5-10). Although *pol* is a single gene, *ts* mutants lie in two complementation groups (4).

It is now known that several other HSV specified proteins also participate in the replication of virus DNA. These include an exonuclease (1,11) and a species termed the major DNA binding protein (here referred to as DBP; also called ICP 8) (12-14). The function of the latter is not clear, although genetic studies have shown that it is necessary for DNA replication (15). It may participate in a complex with exonuclease and DNA polymerase (16). In vitro, DBP exhibits a DNA double helix strand separation activity (14). DBP has an estimated molecular weight of about 130,000 (12,13,15). The gene for DBP (here called *dbp*) is located close to the *pol* gene (15,17).

Regions in HSV DNA which act as origins of replication have been characterized. It is thought that there are three origins: one in U_L, between *dbp* and *pol*, termed Ori_L; and one in each copy of the short repeat region (R_S, see Figure 1), termed Ori_S (18,19). The DNA sequence of the latter is known for both HSV type 1 and type 2 (20,21).

The region of HSV DNA from 0.38 to 0.43 map units is therefore of particular importance in DNA replication, since it contains the *dbp* gene, Ori_L and the *pol* gene in close proximity. In this paper we present an 8.4 kb DNA sequence for this region in HSV-1 strain 17, and analyse the sequence in terms of the *pol* and *dbp* genes and their encoded polypeptides, and of Ori_L. In addition, homologues of both genes can be found in the published genome sequence of the distantly related gammaherpesvirus, Epstein-Barr virus (EBV) (22), although the positions and orientations of these differ from the HSV layout.

MATERIALS AND METHODS

(a) Recombinant Plasmids

Plasmids carrying fragments of HSV-1 strain 17 DNA were used for general sequence analysis, as follows: BamHI *r* and BamHI *v*, in pAT153; EcoRI *f* and EcoRI *m*, in pACYC184; KpnI *v*, KpnI *a'* and

KpnI x, cloned into the PstI site of pAT153 by dG and dC tailing.

(b) DNA Sequence Analysis

The major sequence analysis reported here used the M13/chain terminator system (23). In early work, small restriction fragments were cloned and sequenced. Most of the data were obtained using random fragments made by sonicating DNA and cloning into the SmaI site of M13mp8 (24,25). The buffer gradient gel system of Biggin et al. (26) was used, except that gels contained additional urea, to 9M. Sequence artefacts were resolved by substitution of dITP for dGTP in sequence polymerization reactions, and separately by the use of 6% acrylamide, 1xTBE sequencing gels in an apparatus with a jacket through which water at 80-85°C was pumped.

There exists in HSV-1 BamHI v a region which deletes on cloning into plasmids (see Results), and the sequence of this was determined using virion DNA, as follows. A hexadecanucleotide with sequence from adjacent to the target region was synthesized by solid phase phosphotriester methods (27), using apparatus and reagents from Cruachan Chemical Co., Livingston, Scotland. Genomic HSV DNA was digested with BamHI, and BamHI v fragment isolated. The hexadecanucleotide was annealed to this DNA after strand separation, and this preparation was then used as a primer/template in the chain terminator system.

(c) mRNA Mapping

Termini of HSV mRNAs were mapped using end-labelled DNA restriction fragments or uniformly labelled single stranded probes derived from M13 clones, and S1 nuclease (28). HSV mRNA preparations were the gift of F.J. Rixon.

(d) Computing

A DEC PDP 11/44 under RSX11M was used. DNA sequence data were collated and presented with the database system of Staden (29). Codon usages were evaluated by the program of Staden and McLachlan (30). Amino acid sequence homologies were evaluated using: (1) a matrix comparison program, scoring for identical amino acids (31); (2) a matrix comparison program scoring for related amino acids (32); and (3) an alignment optimising program (33).

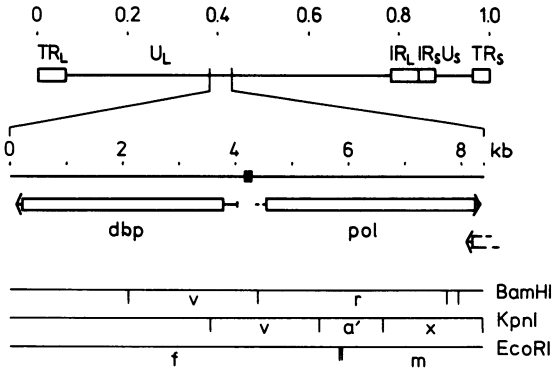


Figure 1. The dbp/pol region in the genome of HSV-1. The upper part of the figure depicts the prototype HSV-1 genome. Large scale repeat elements are shown as open boxes (TR_L and IR_L, TR_S and IR_S) and unique sequences as solid lines (U_L and U_S). The scale indicates fractional genome length map units (total length is about 155,000 bp). The lower part of the figure expands the dbp/pol region; numbering, in bp, corresponds to Figure 2. Dbp and pol mRNAs are indicated, with coding regions as open boxes. Probable mRNA arrangement for the 3' portion of the next gene to the right of pol is also shown. A large palindrome in the locality of Ori_L (see Results) is shown as a filled box. BamHI, KpnI and EcoRI fragments used in sequence analysis are indicated.

RESULTS

(a) Sequence analysis of the DNA polymerase and DBP genes

Figure 1 shows a representation of the HSV-1 genome, with the region from approximately 0.38 to 0.43 map units expanded to show organization of the dbp gene, the origin of replication (Ori_L) and the pol gene. Using plasmid cloned fragments, the DNA sequence was obtained for a region running from the left of dbp to right of pol. However it was known that a region in BamHI v invariably suffered deletion on cloning into a plasmid (18; also, N.D. Stow, personal communication). The sequence of the deleted region was therefore obtained by synthesizing an oligonucleotide, d(GCGGTTGGTATATGTA), which represents sequences 5' to the deletion on the rightward 5' to 3' strand (residues 4069 to 4084 in Figure 2), and using this for primer extension sequencing on a virus genomic DNA fragment. It turned out that 148 bp had been deleted from the plasmid clones used for general sequence analysis. The deletion endpoints were between corresponding points in direct repeats of the sequence CCACGCCC, at residues

4124 and 4272, and a large palindromic sequence was removed (see section (d), below). In addition, an unstable "intermediate-deletion" plasmid, obtained from our colleague N.D. Stow, was analysed and found to have a 60 bp deletion, located within the limits of the stable, 148 bp deleted form (see Figure 2). The autoradiograms from extension sequencing with HSV genomic DNA characteristically suffered from high backgrounds but represented the best data attainable for this region.

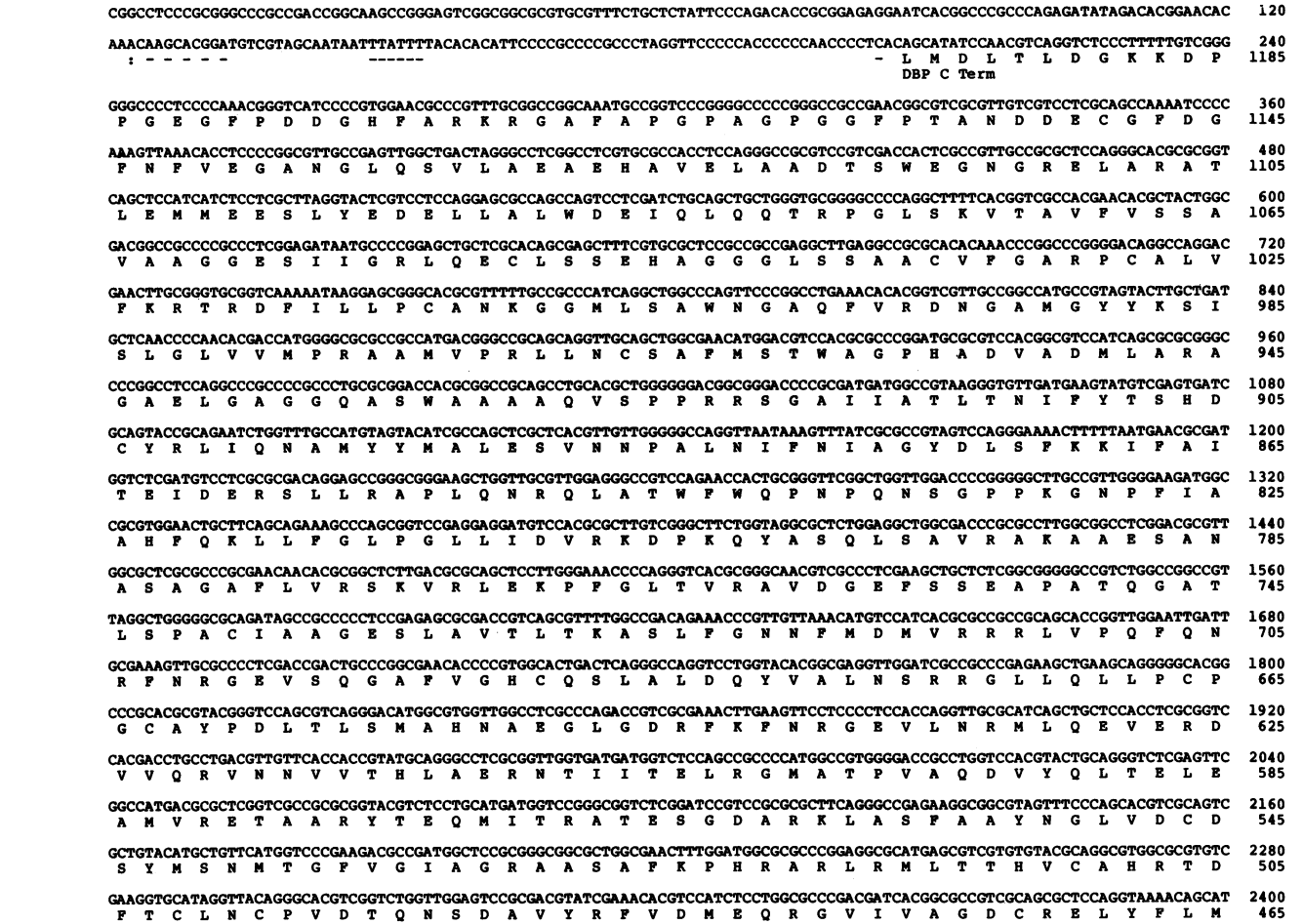
The final sequence obtained, after insertion of the deleting region, is presented in Figure 2 as 8400 residues of 66.1% G+C DNA.

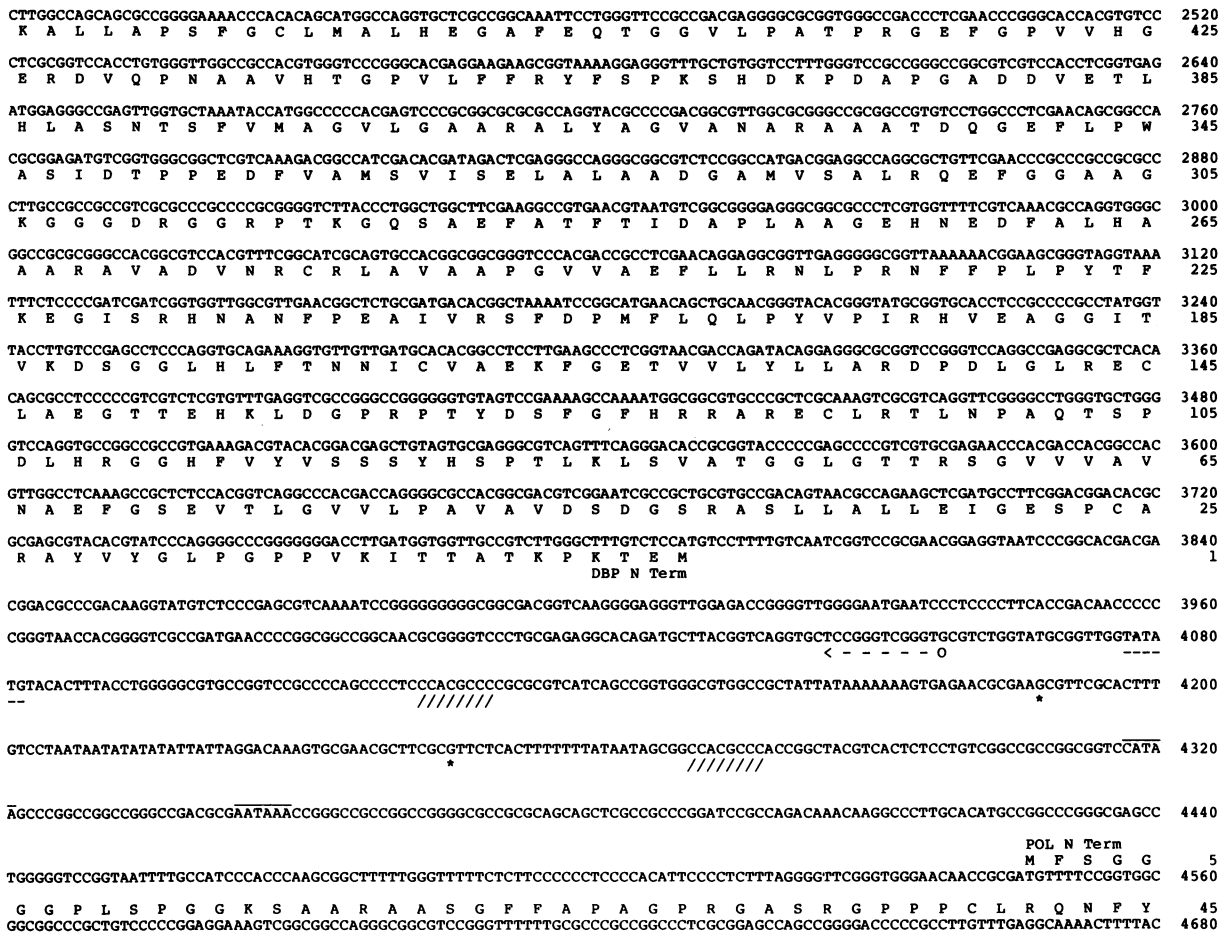
(b) The DNA polymerase gene and protein

Many mutants of HSV DNA polymerase have been shown to map in a genome region of about 5 kb, bounded to the left by the BamHI site between the fragments BamHI v and BamHI r (see refs 9 and 10). This site is at residue 4392 of Figure 2. Holland et al. (34) identified two mRNA species, thought to encode the DNA polymerase, of estimated sizes 4.2 and 4.3 kb, which had their 5' termini close to the v/r BamHI site and were transcribed rightwards. The nature of differences between the two species was not defined.

We attempted to map the 5' termini of pol mRNAs by S1 nuclease analysis with various probes representing sequences near the BamHI site, but did not succeed in detecting a convincing 5' terminus, or any candidate for a splice donor or acceptor site discontinuity. There is thus a possible unresolved complexity in the structure of pol mRNAs' 5' terminal region (see Discussion). In this absence of direct data, and noting that Holland et al. (34), from Northern blot analysis, placed the 5' termini near the v/r BamHI site, we consider that the mRNAs are most likely initiated downstream of the possible TATA consensus sequences, CATAA at 4317 or AATAAA at 4345. The 3' terminus of pol mRNA was mapped approximately to residue 8316 (data not shown), downstream of the polyadenylation signal AATAAA at 8287.

The pol mRNA region so defined contains a 3705 base open reading frame starting with ATG at 4546. Since this reading frame exhibits a typical HSV codon usage and would encode an appropriately sized protein, we conclude that it is the DNA


CGGCCTCCCGCGGGCCCGCCGACCGGCAAGCCGGGAGTCGGCGGCGCGTGCGTTTCTGCTCTATTCCCAGACACCGCGGAGAGGAATCACGGCCCGCCCAGAGATATAGACACGGAACAC 120
AAACAAGCACGGATGTCGTAGCAATAATTTATTTTACACACATTCCCCGCCCCGCCCTAGGTTCCCCCACCCCCCAACCCCTCACAGCATATCCAACGTCAGGTCTCCCTTTTGTCGGG 240
: - - - - - ------ - L M D L T L D G K K D P 1185
DBP C Term
GGGCCCCTCCCCAAACGGGTCATCCCCGTGGAACGCCCGTTTGCGGCCGGCAAATGCCGGTCCCGGGGCCCCCGGGCCGCCGAACGGCGTCGCGTTGTCGTCCTCGCAGCCAAAATCCCC 360
P G E G F P D D G H F A R K R G A F A P G P A G P G G F P T A N D D E C G F D G 1145
AAAGTTAAACACCTCCCCGGCGTTGCCGAGTTGGCTGACTAGGGCCTCGGCCTCGTGCGCCACCTCCAGGGCCGCGTCCGTCGACCACTCGCCGTTGCCGCGCTCCAGGGCACGCGCGGT 480
F N F V E G A N G L Q S V L A E A E H A V E L A A D T S W E G N G R E L A R A T 1105
CAGCTCCATCATCTCCTCGCTTAGGTACTCGTCCTCCAGGAGCGCCAGCCAGTCCTCGATCTGCAGCTGCTGGGTGCGGGGCCCCAGGCTTTTCACGGTCGCCACGAACACGCTACTGGC 600
L E M M E E S L Y E D E L L A L W D E I Q L Q Q T R P G L S K V T A V F V S S A 1065
GACGGCCGCCCCGCCCTCGGAGATAATGCCCCGGAGCTGCTCGCACAGCGAGCTTTCGTGCGCTCCGCCGCCGAGGCTTGAGGCCGCGCACACAAACCCGGCCCGGGGACAGGCCAGGAC 720
V A A G G E S I I G R L Q E C L S S E H A G G G L S S A A C V F G A R P C A L V 1025
GAACTTGCGGGTGCGGTCAAAAATAAGGAGCGGGCACGCGTTTTTGCCGCCCATCAGGCTGGCCCAGTTCCCGGCCTGAAACACACGGTCGTTGCCGGCCATGCCGTAGTACTTGCTGAT 840
F K R T R D F I L L P C A N K G G M L S A W N G A Q F V R D N G A M G Y Y K S I 985
GCTCAACCCCAACACGACCATGGGGCGCGCCGCCATGACGGGCCGCAGCAGGTTGCAGCTGGCGAACATGGACGTCCACGCGCCCGGATGCGCGTCCACGGCGTCCATCAGCGCGCGGGC 960
S L G L V V M P R A A M V P R L L N C S A F M S T W A G P H A D V A D M L A R A 945
CCCGGCCTCCAGGCCCGCCCCGCCCTGCGCGGACCACGCGGCCGCAGCCTGCACGCTGGGGGGACGGCGGGACCCCGCGATGATGGCCGTAAGGGTGTTGATGAAGTATGTCGAGTGATC 1080
G A E L G A G G Q A S W A A A A Q V S P P R R S G A I I A T L T N I F Y T S H D 905
GCAGTACCGCAGAATCTGGTTTGCCATGTAGTACATCGCCAGCTCGCTCACGTTGTTGGGGGCCAGGTTAATAAAGTTTATCGCGCCGTAGTCCAGGGAAAACTTTTTAATGAACGCGAT 1200
C Y R L I Q N A M Y Y M A L E S V N N P A L N I F N I A G Y D L S F K K I F A I 865
GGTCTCGATGTCCTCGCGCGACAGGAGCCGGGCGGGAAGCTGGTTGCGTTGGAGGGCCGTCCAGAACCACTGCGGGTTCGGCTGGTTGGACCCCGGGGGCTTGCCGTTGGGGAAGATGGC 1320
T E I D E R S L L R A P L Q N R Q L A T W F W Q P N P Q N S G P P K G N P F I A 825
CGCGTGGAACTGCTTCAGCAGAAAGCCCAGCGGTCCGAGGAGGATGTCCACGCGCTTGTCGGGCTTCTGGTAGGCGCTCTGGAGGCTGGCGACCCGCGCCTTGGCGGCCTCGGACGCGTT 1440
A H F Q K L L F G L P G L L I D V R K D P K Q Y A S Q L S A V R A K A A E S A N 785
GGCGCTCGCGCCCGCGAACAACACGCGGCTCTTGACGCGCAGCTCCTTGGGAAACCCCAGGGTCACGCGGGCAACGTCGCCCTCGAAGCTGCTCTCGGCGGGGGCCGTCTGGCCGGCCGT 1560
A S A G A F L V R S K V R L E K P F G L T V R A V D G E F S S E A P A T Q G A T 745
TAGGCTGGGGGCGCAGATAGCCGCCCCCTCCGAGAGCGCGACCGTCAGCGTTTTGGCCGACAGAAACCCGTTGTTAAACATGTCCATCACGCGCCGCCGCAGCACCGGTTGGAATTGATT 1680
L S P A C I A A G E S L A V T L T K A S L F G N N F M D M V R R R L V P Q F Q N 705
GCGAAAGTTGCGCCCCTCGACCGACTGCCCGGCGAACACCCCGTGGCACTGACTCAGGGCCAGGTCCTGGTACACGGCGAGGTTGGATCGCCGCCCGAGAAGCTGAAGCAGGGGGCACGG 1800
R F N R G E V S Q G A F V G H C Q S L A L D Q Y V A L N S R R G L L Q L L P C P 665
CCCGCACGCGTACGGGTCCAGCGTCAGGGACATGGCGTGGTTGGCCTCGCCCAGACCGTCGCGAAACTTGAAGTTCCTCCCCTCCACCAGGTTGCGCATCAGCTGCTCCACCTCGCGGTC 1920
G C A Y P D L T L S M A H N A E G L G D R F K F N R G E V L N R M L Q E V E R D 625
CACGACCTGCCTGACGTTGTTCACCACCGTATGCAGGGCCTCGCGGTTGGTGATGATGGTCTCCAGCCGCCCCATGGCCGTGGGGACCGCCTGGTCCACGTACTGCAGGGTCTCGAGTTC 2040
V V Q R V N N V V T H L A E R N T I I T E L R G M A T P V A Q D V Y Q L T E L E 585
GGCCATGACGCGCTCGGTCGCCGCGCGGTACGTCTCCTGCATGATGGTCCGGGCGGTCTCGGATCCGTCCGCGCGCTTCAGGGCCGAGAAGGCGGCGTAGTTTCCCAGCACGTCGCAGTC 2160
A M V R E T A A R Y T E Q M I T R A T E S G D A R K L A S F A A Y N G L V D C D 545
GCTGTACATGCTGTTCATGGTCCCGAAGACGCCGATGGCTCCGCGGGCGGCGCTGGCGAACTTTGGATGGCGCGCCCGGAGGCGCATGAGCGTCGTGTGTACGCAGGCGTGGCGCGTGTC 2280
S Y M S N M T G F V G I A G R A A S A F K P H R A R L R M L T T H V C A H R T D 505
GAAGGTGCATAGGTTACAGGGCACGTCGGTCTGGTTGGAGTCCGCGACGTATCGAAACACGTCCATCTCCTGGCGCCCGACGATCACGGCGCCGTCGCAGCGCTCCAGGTAAAACAGCAT 2400
F T C L N C P V D T Q N S D A V Y R F V D M E Q R G V I V A G D C R E L Y F L M 465


DBP N Term
POL N Term

```
 N  P  Y  L  A  P  V  G  T  Q  Q  K  P  T  G  P  T  Q  R  H  T  Y  Y  S  E  C  D  E  F  R  F  I  A  P  R  V  L  D  E  D    85
AACCCCTACCTCGCCCCAGTCGGGACGCAACAGAAGCCGACCGGGCCAACCCAGCGCCATACGTACTATAGCGAATGCGATGAATTTCGATTCATCGCCCCGCGGGTGCTGGACGAGGAT  4800

 A  P  P  E  K  R  A  G  V  H  D  G  H  L  K  R  A  P  K  V  Y  C  G  G  D  E  R  D  A  P  P  R  R  V  G  R  L  L  A  A   125
GCCCCCCCGGAGAAGCGCGCCGGGGTGCACGACGGTCACCTCAAGCGCGCCCCCAAGGTGTACTGCGGGGGGGACGAGCGCGACGCTCCTCCGCGTCGGGTCGGGCGGCTTCTGGCCGCG  4920

 A  S  R  L  W  G  G  V  D  H  A  P  A  G  F  N  P  T  V  T  V  F  H  V  Y  D  I  L  E  N  V  E  H  A  Y  G  M  R  A  A   165
GCGTCGCGCCTGTGGGGCGGCGTGGACCACGCCCCGGCGGGGTTCAACCCCACCGTCACCGTCTTTCACGTGTACGACATCCTGGAGAACGTGGAGCACGCGTACGGCATGCGCGCGGCC  5040

 Q  F  H  A  R  F  M  D  A  I  T  P  T  G  T  V  I  T  L  L  G  L  T  P  E  G  H  R  V  A  V  H  V  Y  G  T  R  Q  Y  F   205
CAGTTCCACGCGCGGTTTATGGACGCCATCACACCGACGGGGACCGTCATCACGCTCCTGGGCCTGACTCCGGAAGGCCACCGGGTGGCCGTTCACGTTTACGGCACGCGGCAGTACTTT  5160

 Y  M  N  K  E  E  V  D  R  H  L  Q  C  R  A  P  R  D  L  C  E  R  M  A  A  A  L  R  E  S  P  G  A  S  F  R  G  I  S  A   245
TACATGAACAAGGAGGAGGTCGACAGGCACCTACAATGCCGCGCCCCACGAGATCTCTGCGAGCGCATGGCCGCGGCCCTGCGCGAGTCCCCGGGCGCGTCGTTCCGCGGCATCTCCGCG  5280

 D  H  F  E  A  E  V  V  E  R  T  D  V  Y  Y  Y  E  T  R  P  A  L  F  Y  R  V  Y  V  R  S  G  R  V  L  S  Y  L  C  D  N   285
GACCACTTCGAGGCGGAGGTGGTGGAGCGCACCGACGTGTACTACTACGAGACGCGCCCCGCTCTGTTTTACCGCGTCTACGTCCGAAGCGGGCGTGTGCTGTCGTACCTGTGCGACAAC  5400

 F  C  P  A  I  K  K  Y  E  G  G  V  D  A  T  T  R  F  I  L  D  N  P  G  F  V  T  F  G  W  Y  R  L  K  P  G  R  N  N  T   325
TTCTGCCCGGCCATCAAGAAGTACGAGGGTGGGGTCGACGCCACCACCCGGTTCATCCTGGACAACCCCGGGTTCGTCACCTTCGGCTGGTACCGTCTCAAACCGGGCCGGAACAACACG  5520

 L  A  Q  P  A  A  P  M  A  F  G  T  S  S  D  V  E  F  N  C  T  A  D  N  L  A  I  E  G  G  M  S  D  L  P  A  Y  K  L  M   365
CTAGCCCAGCCGGCGGCCCCGATGGCCTTCGGGACATCCAGCGACGTCGAGTTTAACTGTACGGCGGACAACCTGGCCATCGAGGGGGGCATGAGCGACCTACCGGCATACAAGCTCATG  5640

 C  F  D  I  E  C  K  A  G  G  E  D  E  L  A  F  P  V  A  G  H  P  E  D  L  V  I  Q  I  S  C  L  L  Y  D  L  S  T  T  A   405
TGCTTCGATATCGAATGCAAGGCGGGGGGGGAGGACGAGCTGGCCTTTCCGGTGGCCGGGCACCCGGAGGACCTGGTCATCCAGATATCCTGTCTGCTCTACGACCTGTCCACCACCGCC  5760

 L  E  H  V  L  L  F  S  L  G  S  C  D  L  P  E  S  H  L  N  E  L  A  A  R  G  L  P  T  P  V  V  L  E  F  D  S  E  F  E   445
CTGGAGCACGTCCTCCTGTTTTCGCTCGGTTCCTGCGACCTCCCCGAATCCCACCTGAACGAGCTGGCGGCCAGGGGCCTGCCCACGCCCCGTGGTTCTGGAATTCGACAGCGAATTCGAG  5880

 M  L  L  A  F  M  T  L  V  K  Q  Y  G  P  E  F  V  T  G  Y  N  I  I  N  F  D  W  P  F  L  L  A  K  L  T  D  I  Y  K  V   485
ATGCTGTTGGCCTTCATGACCCTTGTGAAACAGTACGGCCCCGAGTTCGTGACCGGGTACAACATCATCAACTTCGACTGGCCCTTCTTGCTGGCCAAGCTGACGGACATTTACAAGGTC  6000

 P  L  D  G  Y  G  R  M  N  G  R  G  V  F  R  V  W  D  I  G  Q  S  H  F  Q  K  R  S  K  I  K  V  N  G  M  V  N  I  D  M   525
CCCCTGGACGGGTACGGCCGCATGAACGGCCGGGGCGTGTTTCGCGTGTGGGACATAGGCCAGAGCCACTTCCAGAAGCGCAGCAAGATAAAGGTGAACGGCATGGTGAACATCGACATG  6120

 Y  G  I  I  T  D  K  I  K  L  S  S  Y  K  L  N  A  V  A  E  A  V  L  K  D  K  K  K  D  L  S  Y  R  D  I  P  A  Y  Y  A   565
TACGGGATTATAACCGACAAGATCAAGCTCTCGAGCTACAAGCTCAACGCCGTGGCCGAAGCCGTCCTGAAGGACAAGAAGAAGGACCTGAGCTATCGCGACATCCCCGCCTACTACGCC  6240

 A  G  P  A  Q  R  G  V  I  G  E  Y  C  I  Q  D  S  L  L  V  G  Q  L  F  F  K  F  L  P  H  L  E  L  S  A  V  A  R  L  A   605
GCCGGGCCCGCGCAACGCGGGGTGATCGGCGAGTACTGCATACAGGATTCCCTGCTGGTGGGCCAGCTGTTTTTTAAGTTTTGCCCCATCTGGAGCTCTCGGCCGTCGCGCGCTTGGCG  6360

 G  I  N  I  T  R  T  I  Y  D  G  Q  Q  I  R  V  F  T  C  L  L  R  L  A  D  Q  K  G  F  I  L  P  D  T  Q  G  R  F  R  G   645
GGTATTAACATCACCCGCACCATCTACGACGGCCAGCAGATCCGCGTCTTTACGTGCCTGCTGCGCCTGGCCGACCAGAAGGGCTTTATTCTGCCGGACACCCAGGGGCGATTTAGGGGC  6480

 A  G  G  E  A  P  K  R  P  A  A  A  R  E  D  E  E  R  P  E  E  E  G  E  D  E  D  E  R  E  E  G  G  G  E  R  E  P  E  G   685
GCCGGGGGGGAGGCGCCCAAGCGTCCGGCCGCAGCCCGGGAGGACGAGGAGCGGCCAGAGGAGGAGGGGGAGGACGAGGACGAACGCGAGGAGGGCGGGGGCGAGCGGGAGCCGGAGGGC  6600

 A  R  E  T  A  G  R  H  V  G  Y  Q  G  A  R  V  L  D  P  T  S  G  F  H  V  N  P  V  V  V  F  D  F  A  S  L  Y  P  S  I   725
GCGCGGGAGACCGCCGGCAGGCACGTGGGGTACCAGGGGGCCAGGGTCCTTGACCCCACTTCCGGGTTTCACGTGAACCCCGTGGTGGTGTTCGACTTTGCCAGCCTGTACCCCAGCATC  6720

 I  Q  A  H  N  L  C  F  S  T  L  S  L  R  A  D  A  V  A  H  L  E  A  G  K  D  Y  L  E  I  E  V  G  G  R  R  L  F  F  V   765
ATCCAGGCCCACAACCTGTGCTTCAGCACGCTCTCCCTGAGGGCCGACGCAGTGGCGCACCTGGAGGCGGGCAAGGACTACCTGGAGATCGAGGTGGGGGGGCGACGGCTGTTCTTCGTC  6840

 K  A  H  V  R  E  S  L  L  S  I  L  L  R  D  W  L  A  M  R  K  Q  I  R  S  R  I  P  Q  S  S  P  E  E  A  V  L  L  D  K   805
AAGGCTCACGTGCGAGAGAGCCTCCTCAGCATCCTCCTGCGGGACTGGCTCGCCATGCGAAAGCAGATCCGCTCGCGGATTCCCCAGAGCAGCCCCGAGGAGGCCGTGCTCCTGGACAAG  6960

 Q  Q  A  A  I  K  V  V  C  N  S  V  Y  G  F  T  G  V  Q  H  G  L  L  P  C  L  H  V  A  A  T  V  T  T  I  G  R  E  M  L   845
CAGCAGGCCGCCATCAAGGTCGTGTGTAACTCGGTGTACGGGTTCACGGGAGTGCAGCACGGACTCCTGCCGTGCCTGCACGTTGCCGCGACGGTGACGACCATCGGCCGCGAGATGCTG  7080
```

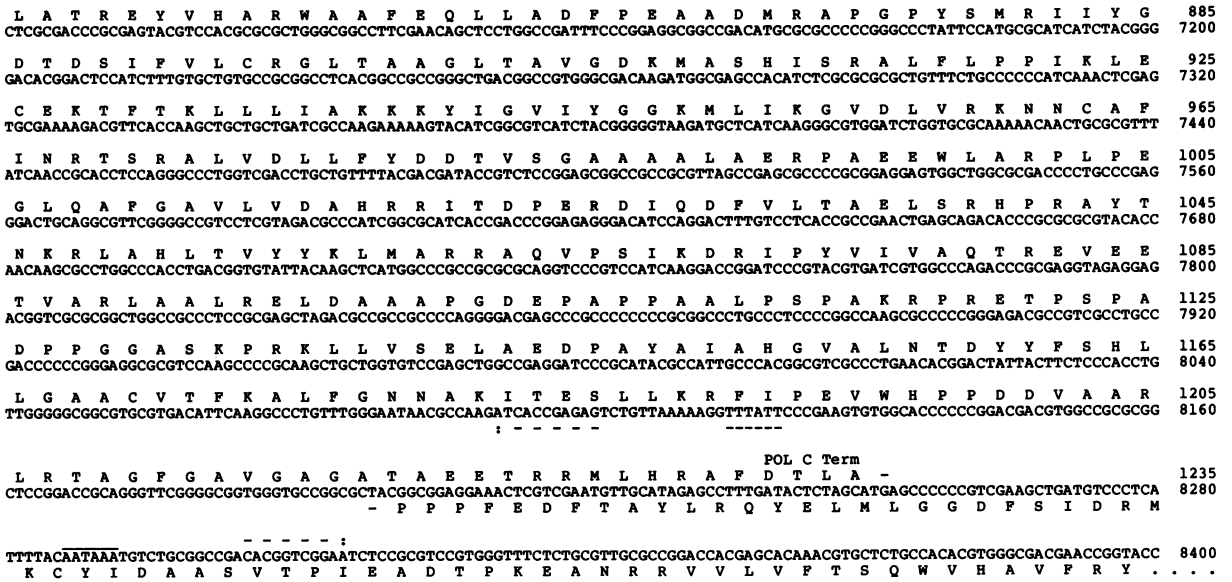
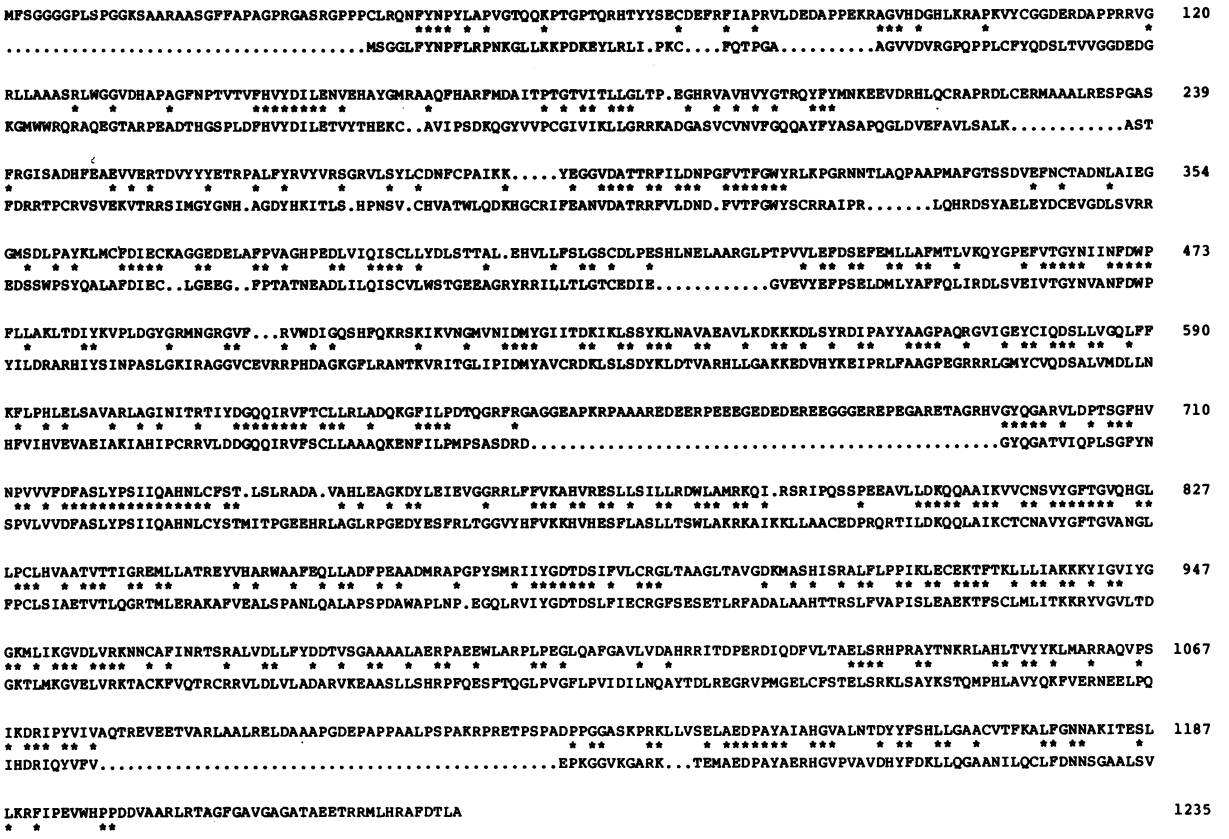


Figure 2. DNA sequence of the dbp/pol region of HSV-1. The sequence is shown corresponding to the rightward 5' to 3' strand only of the dbp/pol region as in Figure 1. The 5' terminus of dbp mRNA is shown as <- - - - - O, and the 3' termini of mRNAs as - - - - - :. Putative TATA boxes and polyadenylation sequences are marked -------. Predicted amino acid sequences are given in single letter code. All symbols for pol (transcribed rightwards) are above the DNA sequence. For dbp and the gene fragment right of pol (both transcribed leftwards) symbols are below the DNA sequence. The limits of the 60 bp "intermediate" plasmid cloning deletion are marked as "*". The direct repeats flanking the major 148 bp plasmid cloning deletion are marked ////////.


MFSGGGGPLSPGGKSAARAASGFFAPAGPRGASRGPPPCLRQNFYNPYLAPVGTQQKPTGPTQRHTYYSECDEFRFIAPRVLDEDAPPEKRAGVHDGHLKRAPKVYCGGDERDAPPRRVG 120
.....MSGGLFYNPFLRPNKGLLKKPDKEYLRLI.PKC....FQTPGA.....AGVVDVRGPQPPLCFYQDSLTVVGGDEDG
RLLAAASRLWGGVDHAPAGFNPTVTVFHVYDILENVEHAYGMRAAQFHARFMDAITPTGTVITLLGLTP.EGHRVAVHVYGTRQYFYMNKEEVDRHLQCRAPRDLCERMAAALRESPGAS 239
KGMWWRQRAQEGTARPEADTHGSPLDFHVYDILETVYTHEKC..AVIPSDKQGYVVPCGIVIKLLGRRKADGASVCVNVFGQQAYFYASAPQGLDVEFAVLSALK.....AST
FRGISADHFEAEVVERTDVYYYETRPALFYRVYVRSGRVLSYLCDNFCPAIKK.....YEGGVDATTRFILDNPGFVTFGWYRLKPGRNNTLAQPAAPMAFGTSSDVEFNCTADNLAIEG 354
FDRRTPCRVSVEKVTRRSIMGYGNH.AGDYHKITLS.HPNSV.CHVATWLQDKHGCRIFEANVDATRRFVLDND.FVTFGWYSCRRAIPR.....LQHRDSYAELEYDCEVGDLSVRR
GMSDLPAYKLMCFDIECKAGGEDELAFPVAGHPEDLVIQISCLLYDLSTTAL.EHVLLFSLGSCDLPESHLNELAARGLPTPVVLEFDSEFEMLLAFMTLVKQYGPEFVTGYNIINFDWP 473
EDSSWPSYQALAFDIEC..LGEEG..FPTATNEADLILQISCVLWSTGEEAGRYRRILLTLGTCEDIE.....GVEVYEFPSELDMLYAFFQLIRDLSVEIVTGYNVANFDWP
FLLAKLTDIYKVPLDGYGRMNGRGVF...RVWDIGQSHFQKRSKIKVNGMVNIDMYGIITDKIKLSSYKLNAVAEAVLKDKKDLSYRDIPAYYAAGPAQRGVIGEYCIQDSLLVGQLFF 590
YILDRARHIYSINPASLGKIRAGGVCEVRRPHDAGKGFLRANTKVRITGLIPIDMYAVCRDKLSLSDYKLDTVARHLLGAKKEDVHYKEIPRLFAAGPEGRRRLGMYCVQDSALVMDLLN
KFLPHLELSAVARLAGINITRTIYDGQQIRVFTCLLRLADQKGFILPDTQGRFRGAGGEAPKRPAAAREDEERPEEEGEDEDEREEGGGEREPEGARETAGRHVGYQGARVLDPTSGFHV 710
HFVIHVEVAEIAKIAHIPCRRVLDDGQQIRVFSCLLAAAQKENFILPMPSASDRD.....GYQGATVIQPLSGFYN
NPVVVFDFASLYPSIIQAHNLCFST.LSLRADA.VAHLEAGKDYLEIEVGGRRLFFVKAHVRESLLSILLRDWLAMRKQI.RSRIPQSSPEEAVLLDKQQAAIKVVCNSVYGFTGVQHGL 827
SPVLVVDFASLYPSIIQAHNLCYSTMITPGEEHRLAGLRPGEDYESFRLTGGVYHFVKKHVHESFLASLLTSWLAKRKAIKKLLAACEDPRQRTILDKQQLAIKCTCNAVYGFTGVANGL
LPCLHVAATVTTIGREMLLATREYVHARWAAFEQLLADFPEAADMRAPGPYSMRIIYGDTDSIFVLCRGLTAAGLTAVGDKMASHISRALFLPPIKLECEKTFTKLLLIAKKKYIGVIYG 947
FPCLSIAETVTLQGRTMLERAKAFVEALSPANLQALAPSPDAWAPLNP.EGQLRVIYGDTDSLFIECRGFSESETLRFADALAAHTTRSLFVAPISLEAEKTFSCLMLITKKRYVGVLTD
GKMLIKGVDLVRKNNCAFINRTSRALVDLLFYDDTVSGAAAALAERPAEEWLARPLPEGLQAFGAVLVDAHRRITDPERDIQDFVLTAELSRHPRAYTNKRLAHLTVYYKLMARRAQVPS 1067
GKTLMKGVELVRKTACKFVQTRCRRVLDLVLADARVKEAASLLSHRPFQESFTQGLPVGFLPVIDILNQAYTDLREGRVPMGELCFSTELSRKLSAYKSTQMPHLAVYQKFVERNEELPQ
IKDRIPYVIVAQTREVEETVARLAALRELDAAAPGDEPAPPAALPSPAKRPRETPSPADPPGGASKPRKLLVSELAEDPAYAIAHGVALNTDYYFSHLLGAACVTFKALFGNNAKITESL 1187
IHDRIQYVFV.....EPKGGVKGARK...TEMAEDPAYAERHGVPVAVDHYFDKLLQGAANILQCLFDNNSGAALSV
LKRFIPEVWHPPDDVAARLRTAGFGAVGAGATAEETRRMLHRAFDTLA 1235
LQNF...TARPPF.....

polymerase coding region. Again, however, there is a qualification regarding the 5' terminal region of the mRNA, since the first possible ATG is at 4422. This leads into a blocked reading frame and is 124 nucleotides 5' to the assigned translation initiator (see Discussion). The polymerase coding region ends with TGA at 8251, leaving a 63 base 3' non coding region in the mRNA. At residues 8121 to 8116 on the complementary strand is the sequence AATAAA. We believe this to be the mRNA polyadenylation signal for the gene to the right of pol, in a tail to tail arrangement. The two mRNAs would thus overlap by some 220 residues. Further, coding usage analysis suggests that translation in this gene terminates with TAG at 8200 to 8198, so that there is a 17 codon overlap with the pol coding region.

We conclude, with some qualification regarding the N terminus, that HSV-1 DNA polymerase contains 1235 amino acids, with a molecular weight of 136,272, which is in reasonable agreement with previous size estimates. In 1984, Baer et al. (22) published the complete genome sequence for the gammaherpesvirus EBV. EBV and HSV differ greatly in their biological behaviour and in the structures of their genomes (35). However, using an incomplete version of our pol sequence, amino acid homology was found between HSV DNA polymerase and the polypeptide predicted from a previously unassigned open reading frame (called BALF5) in the EBV genome, and this served to identify the EBV DNA polymerase gene (22). Figure 3 shows our complete polymerase amino acid sequence aligned with the EBV derived sequence. There are clearly conserved regions distributed throughout the sequences, and also a number of addition/deletion changes. Most notably, HSV DNA polymerase is longer at both termini, and internally contains 2 major insertions relative to the EBV protein, of 49 residues (starting

Figure 3. Homology between the HSV-1 polymerase sequence and the EBV BALF5 sequence. An alignment is shown between the predicted amino acid sequences for HSV-1 DNA polymerase and the EBV BALF5 reading frame. The HSV-1 sequence is the upper, and numbering is shown for HSV-1 polymerase only. Identical residues are indicated by asterisks, and introduced "gapping" characters by dots.

```
EBV   749   L R V I Y G D T D S L F I E C R G
              . : . : : : : : : : . : .   : : :
HSV   880   M R I I Y G D T D S I F V L C R G
              . .   . : : : : : : . : :     : :
Ad2   864   L K S V Y G D T D S L F V T E R G
```

Figure 4. A local similarity between herpesvirus DNA polymerases and Ad2 DNA polymerase. 17-amino acid portions from EBV, HSV-1 and Ad2 DNA polymerase sequences are shown, with starting positions indicated for each. For the pairs EBV/HSV and HSV/Ad2, identical residues are marked ":" and similar residues (either hydrophobic or strongly basic) are marked ".".

at amino acid 646) and 48 residues, starting at 1078. The first of these is particularly hydrophilic, with 20 acidic side chain residues and 9 basic residues, so is most likely present as a looped out surface structure in the HSV polymerase molecule.

We also compared the HSV-1 DNA polymerase sequence with other available DNA polymerase sequences. No detectable sequence homology was found with Escherichia coli DNA polymerase I (36) or with bacteriophage T7 DNA polymerase (37). Adenovirus 2 (Ad2) polymerase (38) also exhibited no general homology, but residues 864 to 872 showed an 8 residue out of 9 match with HSV polymerase residues 884 to 892. HSV, EBV and Ad2 polymerases are all very similar in an extended version of this sequence (residues 880 to 896 of HSV polymerase), as shown in Figure 4. Ad2 polymerase contains 1056 amino acids, and the "conserved" region thus occupies approximately the same relative position in the polypeptide chain as with the herpesvirus polymerases. The possible significance of this observation is discussed below.

(c) The major DNA binding protein gene and amino acid sequence

Holland et al. (34) identified the dbp mRNA as a 4.2 kb species transcribed leftwards, with its 5' terminus between the KpnI site at residue 3556 in Figure 2, and the BamHI site at 4392. We have mapped the 5' terminus of dbp mRNA approximately to residue 4058 by nuclease S1/hybrid analysis (data not shown). Upstream of the 5' terminal locus is a potential TATA box sequence, CATATA, at 4082 to 4077. The first ATG downstream of the 5' terminus, on the leftward 5' to 3' strand, is at residue 3792. This leads into an 1196 codon open reading frame, which shows a typical HSV codon usage, and which ends at TGA, residues 204 to 202. We conclude that this reading frame encodes the DBP,

giving a protein of molecular weight 128,341. We have not mapped the 3' terminus of dbp mRNA, but presume it to be located near residue 123, downstream of the polyadenylation associated sequence AATAAA at 153 to 148. Dbp mRNA would thus be 3936 residues long, excluding poly(A), with a 5' non coding region of 266 bases, a coding region of 3591 bases (including stop codon) and a 3' non coding region of 79 bases.

The amino acid sequence of DBP shows a near equality of acidic and basic side chain residues, and is clearly not a histone-like species. Aside from histones, the best characterized DNA binding proteins are the prokaryotic site specific class, repressors and others, which bind DNA through side chains in alpha-helical regions (39). DBP does not contain any sequences strikingly similar to those involved in DNA binding with this class. However, a number of hydrophilic side chain residue types are involved in DNA binding by the prokaryotic proteins, and many parts of the DBP sequence are rich in these residues, although no one region emerges as a primary candidate for a DNA binding site.

Acting on a suggestion of Dr. Hans Wolf, Pettenkofer Institut, Munich, we have found that the amino acid sequence of DBP is homologous to that encoded by a previously unassigned open reading frame, BALF2, in the genome of EBV (22). The two sequences are shown aligned in Figure 5. The homology is lower than that seen with the polymerase comparison, and alignment requires introduction of 35 small internal gaps. However, similarity is apparent throughout the sequences, and the positions of the N and C termini correspond closely. We conclude that BALF2 represents EBV's dbp gene.

(d) Origin of DNA replication

It is probable that there are three origins of DNA replication in the HSV genome: Ori_L, which comprises part of the sequence reported in this paper, and two copies of Ori_S, located in the short repeat regions (18,19). Because of the propensity of a sequence comprising all or part of Ori_L to undergo deletion when cloned into E. coli, Ori_L is much less well characterized than Ori_S. Studies on Ori_S in HSV-1 have shown that it can be localized to a 90 bp DNA region (19). This contains an imperfect

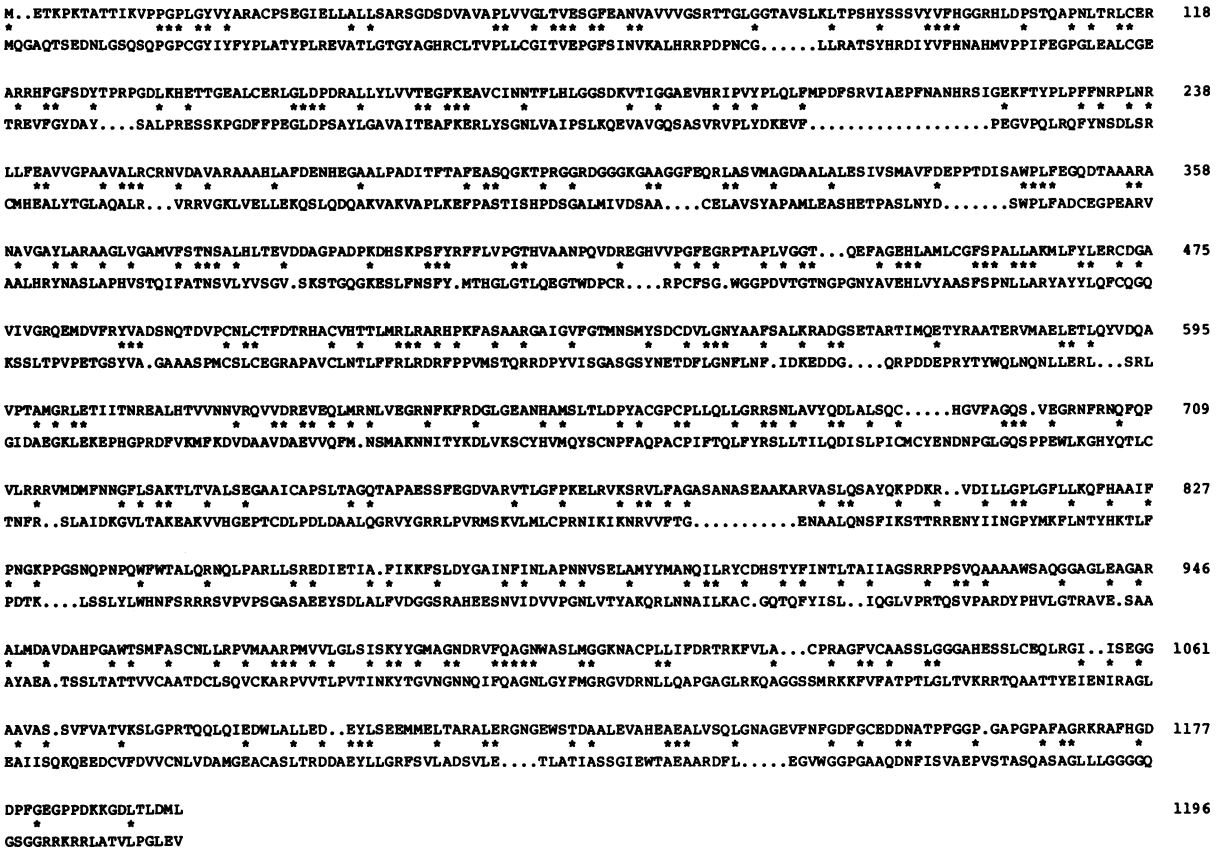


Figure 5. Homology between the HSV-1 DBP sequence and the EBV BALF2 sequence. Layout is as for Figure 3.

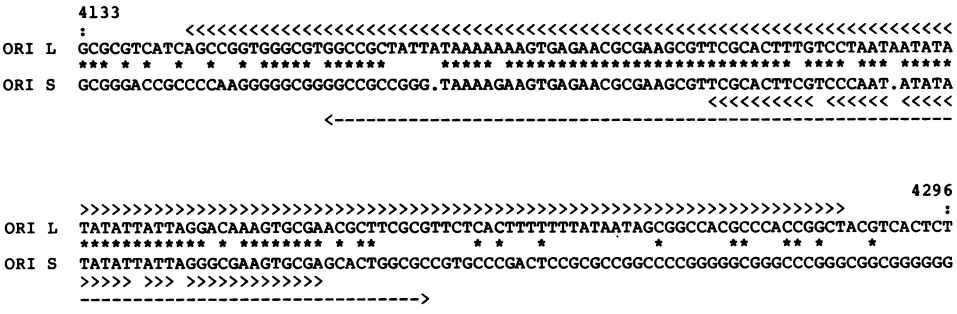


Figure 6. Comparison of HSV-1 DNA replication origins. 164 residues from the DNA sequence of Figure 2 are shown (residues 4143 to 4296), labelled Ori_L, and are compared with the Ori_S sequence (19,20). Identical residues are indicated by asterisks. Two single residue gaps, shown as dots, have been introduced into the Ori_S sequence to maintain alignment. For both sequences, residues in the left arms of palindromes are marked as "<", and residues in right arms as ">". The mapped limits of Ori_S function are shown as <--- ---> (19).

palindromic sequence, with each arm of the palindrome consisting of 21 residues (20). The central part of the palindrome contains the sequence $(AT)_6$, which will be particularly susceptible to strand separation. In HSV-2, strain HG52, sequence analysis demonstrated that each Ori_S is "double", in that two tandem copies of the above sequence are present (21). Gray et al. (40) determined the sequence of the Ori_L region for an HSV-1 strain Angelotti defective species, and found that it contained two copies of a palindrome (like HSV-2 Ori_S), that the palindromes were longer than for Ori_S, and that their sequences were very similar to those in Ori_S. Recently, Weller et al. (41) analysed the sequence of the Ori_L region of HSV-1 strain KOS. In this case a single, long palindromic sequence is present.

As described in section (a), above, the dbp-pol intergenic region of HSV-1 strain 17 DNA undergoes a deletion of 148 residues upon plasmid cloning. In virus genomic DNA this locality contains a palindromic sequence, from 4143 to 4286 in Figure 2. This palindrome is identical to that described for HSV-1 KOS (41). As shown in Figure 6 there is extensive sequence similarity with HSV-1 strain 17 Ori_S, and we presume that all or part of the palindrome is involved in Ori_L function.

Residues 4143 to 4286 comprise a perfect palindrome with arms of 72 residues. Like the Ori_S palindrome, there is a repeated (AT) sequence at the centre of the palindrome. The Ori_L palindrome sequence also contains internal symmetries, as discussed by Weller et al. (41). Comparison with the Ori_S region shows that, first, the whole of the Ori_S palindrome is highly similar to the central region of the Ori_L palindrome, and, second, that the sequence similarity extends beyond the Ori_S palindrome on one side only. This second region is still within the Ori_L palindrome, and in Ori_S lies within the mapped limits for origin function (see Figure 6).

DISCUSSION

The region of HSV-1 DNA described in this paper contains an origin of DNA replication flanked by divergently transcribed genes. This is a commonly found layout, also observed, for instance, with Ori_S of HSV (19,20) and with papovavirus genomes (42). However, the significance of this arrangement in the case of Ori_L, and of the fact that the flanking genes encode proteins of major importance in replication of the DNA, is presently obscure. Comparisons between the sequences of the Ori_L region and of the better characterized Ori_S reveal close similarities, extending over almost the whole of the presently mapped minimal functional unit of Ori_S. The two sequences differ in that the Ori_L region contains a longer palindrome, which is perfect; these characteristics evidently result in severe instability on plasmid cloning. The differences between the two origin sequences appear general in HSV-1 strains, and might be thought to represent some functional difference between the two origin types. An attractive possibility, based on the possession of extended double strand symmetry in Ori_L, is that replication initiation might be bidirectional from Ori_L, but asymmetric from Ori_S. No evidence relating to this is available at present. Another presently obscure question concerns the nature of evolutionary relations between the closely similar DNA sequences in the two origin types.

We did not succeed in definitively mapping a 5' terminus for pol mRNA species, and it is thus possible that the pol mRNA 5'

terminal region could be complex, perhaps involving splicing. In this connection we note that Holland et al. (34) observed two pol mRNA species of similar but distinct gel mobilities. Analysis is complicated by the fact that the 5' terminus must lie close to or within the deleting region of the DNA. We consider that resolution of this problem may require cDNA cloning and sequencing of pol mRNA. Regarding our present interpretation of pol gene structure, the major consequence is that we cannot be certain that the N terminus of the polymerase has been correctly assigned. Examination of the EBV homologue is not helpful since, first, transcription initiation is also obscure in the EBV case and, second, the N termini of the two predicted amino acid sequences are not similar.

Apart from the terminal regions of the HSV-1 DNA polymerase, homology with the EBV BALF5 reading frame is extensive, and provides a convincing identification of BALF5 as encoding EBV DNA polymerase. The homology between these two sequences consists of localities of strongly similar sequence separated by poorly conserved regions and by additions or deletions. We think that at least some of the strongly conserved regions will be found to correspond to structures of particular functional importance, including substrate binding and active site elements of the enzyme. However, we cannot at present identify these: although mutations of varying phenotype (ts and altered drug sensitivity mutants) have been mapped to distinct regions of the gene, the genetic analysis is not yet fine enough to correlate with features of the amino acid sequence (9,10).

The only other eukaryotic DNA polymerase sequence available to us is that of adenovirus 2, which exhibits a striking similarity to HSV-1 and EBV polymerases, in one part of the sequence only (Figure 4). We think it unlikely that this local similarity is adventitious, and consider that it probably represents a common functional element in the polymerase species. However, its evolutionary antecedents are unclear. Ad2 DNA polymerase shows no other sequence similarity of this order with the two herpesvirus polymerases, although there are other, low level correspondences near the highly similar region. We thus face two alternative possibilities, neither of which appears very

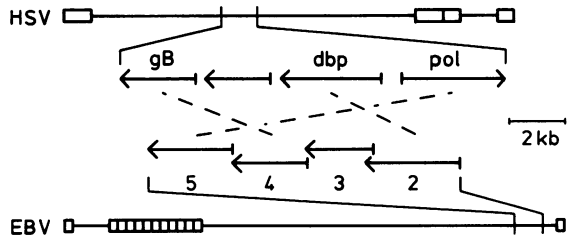


Figure 7. Comparison of arrangements of the dbp/pol region in the HSV-1 and EBV genomes. The upper part of the figure shows the HSV-1 genome, as in Figure 1, with expansion of a region containing dbp and pol and extended to the left to include also the gB gene and another gene between the gB gene and dbp (34, 43; also unpublished data of D.J. McGeoch). The lower part shows the EBV genome, with expansion of a region containing reading frames BALF2, BALF3, BALF4 and BALF5 (22). Arrows are drawn to represent position, size and orientation of the polypeptide coding regions of the various genes. Scale in the expanded regions is indicated by a 2 kb bar. Three pairs of genes with clearly homologous amino acid sequences are joined by dashed lines.

compelling. The first of these is that adenovirus and herpesvirus polymerases do have a common evolutionary origin in whole or part, but have diverged to such an extent that only this region retains strong sequence similarity. This would argue an uniquely critical role for this sequence, unapproached in the remainder of these large proteins, and we regard this as rather unlikely. The other possibility, which we think cannot be excluded at present, is that this is an example of convergent evolution. In this case the uniqueness of the similar region could perhaps best be explained by this sequence being involved in interaction with a conserved host cell factor.

As with the HSV/EBV polymerase homology, the similarity between HSV-1 DBP amino acid sequence and the EBV BALF2 reading frame is clear, although less strong than for the polymerases. Together the homologies provide a view of relations between one region in the genomes of these widely diverged herpesviruses. An additional datum is that homology is readily demonstrable between the amino acid sequence of the HSV-1 glycoprotein gB, whose gene lies two to the left of dbp (43), and the EBV reading frame BALF4, which is the upstream neighbour of the polymerase gene,

BALF5 (22). Two points of large scale dissimilarity in gene layout are evident, as shown in Figure 7. First, the HSV-1 dbp/pol region is located in the interior of the genome, whereas the EBV counterpart is near a terminus of the EBV genome. Second, the relative arrangements of the genes differ in the two viruses. Specifically, pol has been relocated relative to dbp and the gB gene. Thus, during divergence of the present day viruses from a common ancestor, at least two recombinational shuffling events have taken place involving these genes. In addition, it is not thought that the EBV genome possesses an origin of replication in this locality (22).

In conclusion, we expect our analysis of the pol and dbp region of the HSV-1 genome to provide a basis for molecular genetic examination of functions. Additionally, the data provide information on evolutionary relations between parts of the widely diverged HSV and EBV genomes. A copy of this sequence will be deposited with the EMBL Library.

ACKNOWLEDGEMENTS

We owe thanks to Professor J.H. Subak-Sharpe for his support and encouragement and to Dr N.D. Stow for criticism of the manuscript, to Drs A.J. Davison and N.D. Stow for plasmids, and to Dr F.J. Rixon for RNA preparations; also to Drs A.J. Davison, C. Gray, C. Knopf and H. Wolf for access to unpublished data; also to Mr A. Dolan and Mr D. McNab for assistance; also to Cruachan Chemical Co. for the use of an oligonucleotide synthesizer. During the course of this work J.P.Q was supported by an MRC training fellowship.

*Member of the MRC Virology Unit; to whom correspondence should be addressed

+Present address: Laboratory of Molecular Genetics, NICHD, National Institutes of Health, Bethesda, MD 20205, USA

REFERENCES

1. Keir,H.M. and Gold,E. (1963) Biochim. Biophys. Acta 72, 263-276.
2. Purifoy,D.J.M., Lewis,R.B. and Powell,K.L. (1977) Nature 269, 621-623.
3. Knopf,K.W. (1979) Eur. J. Biochem. 98, 231-244.

4. Purifoy,D.J.M. and Powell,K.L. (1981) J. Gen. Virol. 54, 219-222.
5. Hay,J. and Subak-Sharpe,J.H. (1976) J. Gen. Virol. 31, 145-148.
6. Chartrand,P., Stow,N.D., Timbury,M.C. and Wilkie,N.M. (1979) J. Virol. 31, 265-276.
7. Crumpacker,C.S., Chartrand,P., Subak-Sharpe,J.H. and Wilkie,N.M. (1980) Virology 105, 171-184.
8. Coen,D.M., Furman,P.A., Gelep,P.T. and Schaffer,P.A. (1982) J. Virol. 41, 909-918.
9. Honess,R.W., Purifoy,D.J.M., Young,D., Gopal,R., Cammack,N. and O'Hare,P. (1984) J. Gen. Virol. 65, 1-17.
10. Coen,D.M., Aschman,D.P., Gelep,P.T., Retondo,M.J., Weller,S.K. and Schaffer,P.A. (1984) J. Virol. 49, 236-247.
11. Hoffmann,P.J. and Cheng,Y.-C. (1978) J. Biol. Chem. 253, 3557-3562.
12. Bayliss,G.J., Marsden,H.S. and Hay,J. (1975) Virology 68, 124-134.
13. Powell,K.L. and Purifoy,D.J.M. (1976) Intervirology 7, 225-239.
14. Powell,K.L., Littler,E. and Purifoy,D.J.M. (1981) J. Virol. 39, 894-902.
15. Conley,A.J., Knipe,D.M., Jones,P.C. and Roizman,B. (1981) J. Virol. 37, 191-206.
16. Littler,E., Purifoy,D., Minson,A. and Powell,K.L. (1983) J. Gen. Virol. 64, 983-995.
17. Weller,S.K., Lee,K.J., Sabourin,D.J. and Schaffer,P.A. (1983) J. Virol. 45, 354-366.
18. Spaete,R.R. and Frenkel,N. (1982) Cell 30, 295-304.
19. Stow,N.D. and McMonagle,E.C. (1983) Virology 130, 427-438.
20. Murchie,M.-J. and McGeoch,D.J. (1982) J. Gen. Virol. 62, 1-15.
21. Whitton,J.L. and Clements,J.B. (1984) Nucl. Acids Res. 12, 2061-2079.
22. Baer,R., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hatfull,G., Hudson,G.S., Satchwell,S.C., Seguin,C., Tuffnell,P.S. and Barrell,B.G. (1984) Nature 310, 207-211.
23. Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) J. Mol. Biol. 143, 161-178.
24. Deininger,P.L. (1983) Anal. Biochem. 129, 216-223.
25. Messing,J. and Vieira,J. (1982) Gene 19, 269-276.
26. Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
27. Gait,M.J., Matthes,H.W.D., Singh,M., Sproat,B.S. and Titmas,R.C. (1982) Nucl. Acids Res. 10, 6243-6254.
28. Berk,A.J. and Sharp,P.A. (1977) Cell 12, 721-732.
29. Staden,R. (1982) Nucl. Acids Res. 10, 141-156.
30. Staden,R. and McLachlan,A.D. (1982) Nucl. Acids Res. 10, 141-156.
31. Pustell,J. and Kafatos,F.C. (1982) Nucl. Acids Res. 10, 4765-4782.
32. Dayhoff,M.O., Baker,W.C. and Hunt,L.T. (1983) In Methods in Enzymology, Hirs,C.H.W. and Timasheff,S.N. Eds., Vol 91, pp. 524-545, Academic Press, New York.
33. Taylor,P. (1984) Nucl. Acids Res. 12, 447-456.

34. Holland,L.E., Sandri-Goldin,R.M., Goldin,A.L., Glorioso,J.C. and Levine,M. (1984) J. Virol. 49, 947-959.
35. Roizman,B. (1982) In The Herpesviruses, Roizman,B. Ed.,Vol 1, pp. 1-23. Plenum Press, New York and London.
36. Joyce,C., Kelly,W.S. and Grindley,N.D.F. (1982) J. Biol. Chem. 257, 1958-1964.
37. Dunn,J.J. and Studier,F.W. (1983) J. Mol. Biol. 166, 477-535.
38. Gingeras,T.R., Sciaky,D., Gelinas,R.E., Bing-Dong,J., Yen,C.E., Kelly,M.M., Bullock,P.A., Parsons,B.L., O'Neill,K.E. and Roberts,R.J. (1982) J. Biol. Chem. 257, 13475-13491.
39. Ohlendorf,D.H. and Matthews,B.W. (1983) Annu. Rev. Biophys. Bioeng. 12, 259-284.
40. Gray,C.P. and Kaerner,H.C. (1984) J. Gen. Virol. 65, 2109-2119.
41. Weller,S.K., Spadaro,A., Schaffer,J.E., Murray,A.M., Maxam,A.M. and Schaffer,P.A. (1985) Mol. Cell. Biol. 5, 930-942.
42. Bergsma,D.J., Olive,D.M., Hartzell,S.W. and Subramanian,K.N. (1982) Proc. Natl. Acad. Sci. USA 79, 381-385.
43. Bzik,D.J., Fox,B.A., DeLuca,N.A. and Person,S. (1984) Virology 133, 301-314.